2011

Assessment of the Implementation of the Safer Social Networking Principles for the EU on 9 services: Summary Report

August, 2011

By request of the European Commission under the Safer Internet Programme

European Commission
Information Society and Media

August 2011

Please cite as follows:
Donoso, V. (2011). Assessment of the implementation of the Safer
Social Networking Principles for the EU on 9 services: Summary Report.
European Commission, Safer Internet Programme, Luxembourg.


Dailymotion


Google


Microsoft


Skyrock


stardoll
your paperdoll heaven


sulake


YAHOO!
EUROPE

TABLE OF CONTENTS

TABLE OF FIGURES

EXECUTIVE SUMMARY

- This report is part of the European Commission`s commitment to support the industry self-regulatory initiative - the "Safer Social Networking Principles" (also referred to as "the Principles" in this report) signed by 21 social networking companies to date. The report summarizes the findings of the 2nd assessment (Phase B) where 9 social networking websites (SNS) were tested. It also summarizes the main findings related to the analysis of the corresponding self-declarations submitted by the signatories of the Principles involved in this Phase. In the individual self-declarations, providers explain how safety measures have been implemented on their websites.

- The second assessment of the Safer Social Networking Principles aims at determining how well the Principles each SNS committed itself to implement have been put into operation on their corresponding services. The methodology of this second assessment varies slightly in relation to the first evaluation carried out in 2009. Instead of testing all the SNS at once, two Phases have been foreseen. In Phase A, 14 typical SNS were assessed[1], while in Phase B (results summarized in this report) different platforms, namely video-sharing platforms, photo-sharing platforms, virtual worlds, gaming platforms and other platforms have been tested.

- This report consists of two parts. The first part is a general analysis of the main findings across the services assessed. The second part is comprised of individual testing reports of each of the 9 services evaluated in Phase B.

- The methodology employed consists of three main parts, namely, (1) an analysis of the self-declarations submitted by the signatories; (2) the testing (from a user perspective) of the services run by the signatories, and (3) the assessment of how satisfactorily each SNS has implemented their individual commitments in relation to the Principles (as expressed in their individual self-declarations) on the services they run.

- The services tested in this Phase were: Dailymotion, Habbo Hotel, Skyrock, Stardoll, Windows Live, Xbox Live (both the online gaming platform and the console), Flickr, Yahoo! Pulse and YouTube. Stardoll and Yahoo! Pulse were assessed for the first time. The other 7 services were included in the 1st Assessment of the Safer Social Networking Principles for the EU (January, 2010). All these services were tested in their main language versions. Habbo Hotel was tested both in the Finnish and in the English (UK) versions of the service.

MAIN RESULTS OF THE ANALYSIS OF THE SELF-DECLARATIONS SUBMITTED BY THE SIGNATORIES

- The first step of this assessment consisted in determining if each individual self-declaration was in line with the Safer Social Networking Principles. This analysis revealed that among the 9 Social Networking Sites evaluated in this Phase, Principle 7 "Assess the means for reviewing illegal or prohibited content/conduct" and Principle 5 "Respond to notifications of illegal content or conduct" were the best evaluated with all services' commitments being assessed as *very satisfactory*. Principles 1 "Raise awareness"and Principle 2 "Age-appropriate services" were assessed as *very satisfactory* in 8 of the 9 services assessed and as *rather satisfactory* in only 1 service.

[1] For an overview of the main results for Phase A, please visit: http://ec.europa.eu/information_society/activities/social_networking/eu_action/implementation_princip_2011/index_en.htm

- Commitments related to Principle 3 were assessed as *very satisfactory* in 5 services' self-declarations, as *rather satisfactory* in 3 and as *unsatisfactory* in 1. One of the main weaknesses observed in the self-declarations is that services do not make explicit in their individual commitments if (and/or how) they ensure that the default full profiles of those registering under the age of 18 have been set to 'private'. It is important to point out, however, that platforms tested in Phase B are not "typical" social networking sites. This means that in several services user profiles are not the main point of entry or of interest for a user. This is specially the case of photo-sharing and video-sharing platforms where users would typically visit a photostream or a video channel and not necessarily a user's profile.

- Another important point observed in the analysis of Principle 3 is that several self-declarations do not make clear which information from minors' profiles is searchable or visible by default to other users beyond the minors' approved lists of contacts. However, some services do make explicit that the profiles in their services are supposed to contain minimal personal information or that very little personal information (sometimes only username and e-mail) is required to open an account. Even in these cases, it is not clear if additional information added by minors to their user profiles would be mapped to the user profile or made visible to users beyond the minor's approved "friends".

- Commitments expressed in the self-declarations regarding Principle 4 "Easy to use mechanisms for reporting violations" were assessed as *very satisfactory* in only 2 self-declarations and as *rather satisfactory* in 7. The main weakness has to do with the lack of explicit information (in their self-declarations) regarding the age-appropriateness, availability and user-friendliness of the reporting mechanisms available on the services and the lack of information regarding whether user reports were acknowledged or if users were provided with some indication of how their reports were typically handled.

- Only 2 services were *very satisfactory* on all the Principles assessed in their self-declarations.

MAIN RESULTS OF THE TESTS ON THE WEBSITES

- Following the analysis of the self-declarations, each service was tested. The results of these tests were analysed in order to determine the extent to which each SNS had implemented the commitments from their self-declarations on their services. This analysis shows that Principle 1 "Raise awareness" was the best assessed in terms of the implementation of the provider's commitments expressed in their self-declarations with all the services tested being evaluated as *very satisfactory*. Indeed, all the SNSs evaluated provided safety information, guidance and/or educational materials for minors, parents, teachers and/or carers and most of the times this information was *age-appropriate*, *easy-to-understand* and *easy to find* as indicated in the individual self-declarations.

- In relation to Principle 2 "Age appropriate services", 7 services were assessed as *very satisfactory* as regards the implementation of their self-declarations on their services while 2 services were evaluated as *rather satisfactory*. No service was evaluated as *unsatisfactory* regarding this Principle. The best evaluated services where those where, as expressed in their individual self-declarations, signing-up for underage users was denied, where effective mechanisms to prevent re-registration as expressed in their self-declarations were in place and where minors were not confronted with any type of inappropriate content.

  In terms of the implementation of the services' commitments expressed in the self-declarations on the websites tested, Principle 3 "Empower users" was assessed as *very satisfactory* in only 5 services and as *rather satisfactory* in 4. Nevertheless, it is worthwhile mentioning that no service was assessed as *unsatisfactory* on the websites tested. Services that effectively implemented their commitment to empower users were those who stated in their self-declarations that they would limit the visibility of minors' personal information to users beyond the approved contact list and that ensured that minors could not be contacted by users outside their friends and who effectively did so in their services.

- As regards Principle 4, 8 of the 9 SNS providers were assessed as *very satisfactory* and only 1 as *rather satisfactory*. Services that were assessed as *very satisfactory* committed themselves in their self-declarations and effectively provided reporting mechanisms on their services that were both user-friendly and effective. Furthermore, they reacted promptly to the minor's report and effective action (in line with what was stated in the provider's self-declaration) was taken. Even though several services did not self-declare that the reporting mechanisms available on their services were child-friendly or effective, testing on the services themselves demonstrated that in most of the cases, these mechanisms were, indeed, user-friendly and efficient.

- Principle 6 "Encourage safe use approach to privacy" was assessed as *very satisfactory* in 7 services and as *rather satisfactory* in 2 services. The best assessed services are those that committed themselves in their self-declarations and effectively offered accessible and at all times available privacy settings on their services that allow users to have a finer degree of control over which aspects of a user's information are visible and the kinds of online communication that are allowed on these services.

- 2 services implemented the commitments expressed in their individual self-declaration very satisfactorily on the 5 Principles [2] tested on their websites. Only one of these services was also very satisfactorily assessed in all the Principles in its self-declaration.

SUMMARY PRINCIPLE 1 – "RAISE AWARENESS"

- Overall, the assessment of Principle 1 is very positive on the website and consistent with the providers` commitments in their self-declarations. All the SNS assessed were evaluated as *very satisfactory* on their services.

- In terms of the availability and easiness of the safety information on the websites, 8 services include in their services safety information, guidance and/or educational materials specifically targeted at children. In all these services this information was easy for minors to understand and in 7 services it was also easy to find. Only in one of the services the information provided was not specifically targeted at minors, but rather at their parents or guardians. Nevertheless, the service in question provides plenty of general safety information and tips throughout the site which could also be easily understood by minors.

- All the services provide safety information, guidance and/or educational materials on their websites specifically targeted at parents, guardians and/or teachers. In all the cases this information was easy for parents to understand and in 8 it was also easy to find.

- Regarding the type of safety information provided, all the services include general safety information. 8 services include specific information about pornography or sexual content, bullying, hate speech and risks associated to divulging personal information on their services. 6 services provide information on the risks associated to posting sexually provocative pictures and 5 include information on violence and adults with sexual interest on children. Information on other topics such as self-harm (suicide, anorexia, bulimia, etc.) is less common. Safety information is usually presented in the form of written texts, external links or referrals to (educational) organizations active in child safety and concrete examples related to safety. In 6 services, this information is also provided via audio-visual fragments and 1 service presented this information via games.

---

[2] Because of ethical reasons Principles 5 and 7 were not tested on the platforms themselves but only in the services' self-declarations. Therefore even though the services'self-declarations were assessed in terms of the 7 Safer Social Networking Principles, the platforms themselves were only tested on 5 Principles.

- All the SNS provide general Terms of Use or Service as well as an adapted, shorter and more child-friendly version of the Terms. This shorter version is usually presented in the form of Community guidelines, User Agreements or House rules. In all the services assessed the adapted version of the Terms of Service was easy-to-understand and easy to find. Both the general versions of the Terms of Use as well as the shorter child-friendly versions thereof include explicit information on what constitutes inappropriate or forbidden behaviour on the service and the consequences thereof. In the majority of the cases this information is presented in the form of written texts, external links or referrals to relevant organizations and/or concrete examples that illustrated the conditions of use of the service. Only in one service, this information was also provided via audio-visual fragments.

SUMMARY PRINCIPLE 2 –"AGE APPROPRIATE SERVICES"

- In terms of the implementation of the services' individual commitments expressed in the self-declarations on the websites tested, Principle 2 was assessed as *very satisfactory* in 7 services and as *rather satisfactory* in 2 services. The best evaluated services where those where services committed themselves and effectively denied sign-up on the service to underage users, where effective mechanisms to prevent re-registration were in place and where minors were not confronted with any type of inappropriate content including advertising (in the case of those services that referred to specific measures related to advertising in their self-declarations).

- 5 of the 9 SNS have set up a minimum age requirement in order for users to be able to sign up to their services. In 1 of these services the minimum age requirement is 12 and in 3 services it is 13 years old. In the service tested in two language versions, the minimum registration age was 10 in the Finnish version and 13 in the UK version of the site. Within the services that are age-restricted the most common age verification mechanism is self-declaration of age.

- In 4 services no minimum age registration applies. However, in one of these services, parental consent is needed to approve a Community Membership of the site for all children under the age of 13. In another non age-restricted service, accounts for under 18 year olds need to be set up by an adult. This is verified by entering valid credit card details. In the latter case, parents are also required to be responsible for child or teen accounts and to authorise any amendments to default settings.

- The 5 services intended to be age-restricted prevent (initial) sign-up by underage users on their sites, although in 2 of these services users could eventually re-register on the site by changing their initial age to one just above the minimum age required by the SNS. In one service cookies had been installed to avoid re-registration, so sign-up was possible only after removing these cookies.

- In 2 services minors could not re-register on the site. It must be noted though that as no 100% reliable age-verification mechanism exists up to date, even in the cases where age-restriction mechanisms proved to be effective, it was possible for underage users to register on the site by creating a completely new user account with an age above the minimum required by the service.

- Practically none of the services tested included "adult" sections or services, except one where an "adult-only" channel (inaccessible to minors by default) displays "erotic" and/or sexually explicit videos. In all the other services tested all services and sections were appropriate for all audiences.

- Regarding the appropriateness of content found on the services, this depends largely on the adequate tagging or labelling of the materials uploaded by users themselves. This is specially the case of photo-sharing and video-sharing platforms where due to the large amount of content uploaded every minute, it is practically impossible for services to review all this information manually. Services, therefore, have implemented community-driven mechanisms such as tagging, labelling and flagging of content to support the classification and age-restriction of materials uploaded to the sites.

- During testing labelling or tagging mechanisms (associated to internal filtering processes) proved to be effective, although not infallible because in a few cases it was possible to identify content that could be considered as inappropriate for younger children or adolescents, including drug paraphernalia, self-harm triggering and highly erotic content (clearly not meant for artistic or educational purposes).

- Although advertising is not included in the Safer Social Networking Principles for the EU, 3 of the 9 services tested explicitly indicate in their self-declarations that advertising on the services they run is always age-appropriate. In these 3 services, no advertising that could be considered as inappropriate for minors was found.

- In two of the services tested, online communication was real-time moderated both through the use of text filtering and by staff. In one of the services, divulging personal information such name of school attended or phone number was blocked. Text filtering was also successfully applied when racist comments or inappropriate language was posted in comments on some of the sample accounts set up. In the other service tested it was possible to obtain some information from the minor (place of residence, school and address), but phone numbers or e-mail addresses were effectively prevented from being exchanged.

- 5 of the services tested employ some form of parental control on their website**.** These mechanisms vary greatly from service to service. While some services provide sophisticated parental control tools that allow parents or guardians to effectively monitors their children's use of the service, other services employ basic control mechanisms such as informing parents that their child has opened an account on the service or asking for parental consent during registration. Only one of the services tested provided a sophisticated, but still user-friendly parental control software. The software was found easy to install and operate and it proved to be effective in providing a high degree of parental control (and monitoring) over all aspects of online search, web access and online communication.

SUMMARY PRINCIPLE 3 – "EMPOWER USERS"

- According to the analysis of the self-declarations, 3 services explicitly state that the default settings for users under 18 years old are set to friends only. Of these three services, only one does not display personal, non-identifiable information from the minor to (non) users beyond the minor`s approved contact list by default. In another one of these services limited personal information was displayed, but it was accessible to friends of friends. In the other 6 self-declarations analysed, it is not completely clear what type of information is visible to other users by default. Testing on the services demonstrated that in only one of these 6 services profiles of minors did not display personal, non-identifiable information to (non) users beyond the minor`s approved contact list by default. Thus, in sum, only 2 services make minors' personal and identifiable information visible by default only to their approved list of contacts. This means that even if minors add extra information about themselves (not required during registration) to their profiles this information is not displayed to users beyond the minor's contacts by default.

- Regarding the information contained in user profiles, several services state in their self-declarations that very little personal information from minors is required at registration. However, some of these self-declarations also state that if users wish so, they may add more personal information to their profiles. It is not clear from the self-declarations, though, if this optional information is (automatically) mapped into the user's profile and, thus, made visible to other users (beyond the minor's approved contact list) by default. Testing demonstrated that a considerable amount of personal information - including information added by users after registration- was displayed to users beyond the minor`s approved contacts list by default. In some cases, the (full) name of the minor, gender, age, the school they attend, the minor's location (although not their address), profile picture, pictures and videos uploaded by the minor or their contacts, comments posted on the minor`s profiles, the minor`s contact list or online status, their interests and/or hobbies, etc. were displayed to either "friends of friends" and/or non-friends. It is important to stress, however, that this information did not always lead to the identification of minors and, therefore, does not necessarily constitute an imminent safety

risk. Furthermore, platforms tested in Phase B are not "typical" social networking sites. This means that in several services user profiles are not the main point of entry or of interest for a user. This is specially the case of photo-sharing and video-sharing platforms where users would typically visit a photostream or a video channel, but not necessarily a user's profile.

- By default, in all the services tested minors could be contacted via friend requests, in 7 services they could be contacted via public messages and in 3 services they could be contacted via private messages by users beyond their approved list of contacts. It must be noted though that in one service, only 17 year old users could be contacted by non-friends by default (including non-registered users) via private messages or by posting comments on the minor's public blog. However, users younger than 17 (12 and 16 year olds in the case of our tests) could only be contacted by friends.

- By default, only in one of the 9 services tested profiles of minors could be found by name searches[3] either via internal or external search engines (e.g. Google, Bing or Yahoo!). Nevertheless, this did not guarantee that the profile of minors could not be found in the other services tested via other mechanisms. For instance, in 6 of the services the profiles of minors could be directly accessed by friends of friends. In these cases, friends of friends have, by default, access to their friend's contact lists and can, thus, get access to profiles of minors who do not belong to their own lists of contacts. In this sense, profiles of minors could be considered as "unlisted", i.e. they do not appear in searches but may be viewed by anyone with a link to the profile. User-generated content is another way for people (including sometimes non-registered users of the sites) to get access to minor's profiles. Once a minor's blog, individual videos, pictures, video channels or photostream are found (e.g. via an external search engine or via a link to such content posted somewhere on the web, sent via e-mail, etc.), access to the minor's profile is granted. In all these cases, the amount of personal information displayed on the profiles depends on the amount of information provided by the minor as well as on the default settings and the additional privacy settings that the minor may have set up.

- In the 9 services tested users could block other users and reject friends 'requests.

SUMMARY PRINCIPLE 4 – "EASY TO USE MECHANISMS FOR REPORTING VIOLATIONS"

- All the services tested provided one or more mechanisms to report inappropriate content or contact on their website including a general report button and/or a report button next to user-generated content (e.g. to flag inappropriate pictures, videos or comments). 7 services provide both. Report forms were available in all the services and reporting via e-mail was also possible in 2 services.

- All the services assessed provide age-appropriate, user-friendly and easily accessible reporting mechanisms. These reporting mechanisms were at all times available. However, this was not always explicitly stated in the individual self-declarations.

- In 8 services (including the one that was tested in two language versions) reports were acknowledged by the provider. 7 services responded to the user reports taking some action such as deleting or age-restricting the flagged content from the site, issuing warnings to the users who had violated the Terms of Use, or guiding the user who filed the complaint on how to solve the problem themselves.

- In 6 services, the reports were responded within 24 hours. In one service a reply to the reporting user was sent within 24 hours, but the reported content was deleted from the site only after 48 hours.

[3] For the sake of testing, profiles of minors were created and these were searched by typing the full name of the minor between inverted commas (e.g. "John Smith") in both the internal search engine of the service tested and external ones such as Google. Following this method, in only 1 service it was possible to find the profile of minors. However, when searching for username (not full name) or for terms included in the names or tags of videos or photos, profiles of minors could be retrieved in some cases.

- In 6 cases the inappropriate reported content was removed from the site or was age-restricted by the service provider and only one service took some kind of action against the offender.

SUMMARY PRINCIPLE 5 – "RESPOND TO NOTIFICATIONS OF ILLEGAL CONTENT/CONDUCT"

- Because of ethical reasons Principle 5 was not tested on the website.

- Principle 5 was evaluated as very satisfactory in all the services in relation to the self-declaration statements. In their self-declaration all the service providers claim that they have effective and expeditious processes in place to review (and eventually remove) content found offending from their services.

- All the service providers assessed claim to have arrangements in place to share reports of illegal content with the corresponding law enforcement bodies.

- Only 3 services explicitly state in their self-declarations that they provide links on their websites to other local agencies and organisations in order to support the reporting of illegal content or conduct on their services.

SUMMARY PRINCIPLE 6 – "ENCOURAGE SAFE USE APPROACH TO PRIVACY"

- All the services assessed offer their users (including minors) a range of privacy settings. These settings enable users to control who may have access to the information contained in their profile. Privacy settings were user-friendly and accessible at all times in all the services assessed.

- Regarding the content posted on one's profile, some services offer users the option to set up privacy settings for individual pieces of content (e.g. videos or pictures).

- Some services provide privacy settings that allow distinctions between different types of users, for instance, "contacts", "friends and family", "VIP friends", "all users", etc. These can be used to allow or restrict access to certain types of content or activities for different users (e.g., deciding who can have access to guestbook, photos, videos or blogs; commenting on someone's profile, etc.).

- The range and modality of privacy options vary from service to service. Some services offer advanced, but still user-friendly privacy options that allow users to decide which individual piece(s) of content to share with specific (groups of) users. Services which offer more advanced settings allow users to modify individual settings with a finer degree of control over which aspects of a user's information are visible and the kinds of online communication that are allowed. Other services offer more limited privacy settings which lack complexity and do not allow users to customize privacy settings regarding specific groups of people or specific content.

- Privacy settings options are accompanied by supporting information in all the services assessed. Supporting information and guidance help users make informed decisions regarding their privacy settings options and/or instruct users on how to set or change their privacy settings.

- In 5 of the 9 services assessed (some of) the information provided by minors during registration is automatically mapped into the user`s profiles. However, only in 3 cases users were made aware that this would happen. Still, in 4 of the services that automatically map information into the user`s profiles it was possible at a later stage to make private or public (most of) this information.

- In 7 of the 9 services tested it was possible for users to delete their profile themselves. One service does not delete profiles on request and in 1 service it is not possible to delete a profile unless the main ID account (which is also used to access all the other services offered by this provider such as e-mail, photo-sharing applications, etc.) is terminated. In the latter case, though, the privacy settings can be

used to hide the profile from all other users. Deletion of a user profile is straightforward and easy to manage in the 7 services that provide this option.

SUMMARY PRINCIPLE 7- "REVIEWING ILLEGAL OR PROHIBITED CONTENT/CONDUCT"

- Because of ethical reasons, Principle 7 was not tested on the website.

- According to the analysis of the self-declarations, Principle 7 was assessed as *very satisfactory* in all the services assessed. The main mechanisms for services to identify potential risks to minors are user-generated reports and the use of automated mechanisms such as employing text or image filters to identify potential threats to minors. According to the 9 self-declaration statements analysed, both mechanisms are available in all the services tested.

- According to the self-declarations, specially trained personnel reviews inappropriate content in all the services analysed. Three service providers state that they also employ human moderators who interact in real-time with children and young people. However, only in 2 of these services the self-declaration explicitly refers to the steps taken to minimize the risk of employing candidates who may be inappropriate for interacting with minors in real-time. These include strict selection procedures prior to hiring moderators and dedicated training specially tailored for them.

INTRODUCTION

In 2008, as part of its Safer Internet Plus Programme, the European Commission gathered 18 of the major online social networks active in Europe as well as researchers and child welfare organizations to form a European Social Networking Task Force to discuss guidelines for the use of social networking sites by children and young people. As a result "the Safer Social Networking Principles for the EU" were developed by social networking service providers in consultation with the Task Force. The aim was to "provide good practice recommendations for the providers of social networking and other user interactive sites, to enhance the safety of children and young people using their services".

The guidelines were adopted voluntarily by the major online social networks active in Europe, and signed on Safer Internet Day on February 10th 2009. The Principles are meant as a guidance to SNS providers when they seek to minimize potential harm to children and young people ("Safer Social Networking Principles of the EU," 2009: 1). They recommend a wide range of good practice approaches, allowing for the diversity and judgment of the social networks themselves in terms of relevance and implementation. Within the context of the Principles, "Social Networking Services" are defined as services that combine the following features ("Safer Social Networking Principles of the EU," 2009: 3):

- A platform that promotes online social interaction between two or more persons for the purposes of friendship, meeting other persons, or information exchange;
- Functionality that lets users create personal profile pages that contain information of their own choosing, such as the name or nickname of the user, photographs placed on the personal page by the user, other personal information about the user, and links to other personal pages on the service of friends or associates of the user that may be accessed by other users or visitors to the service;
- Mechanisms to communicate with other users, such as a message board, electronic mail, or instant messenger; and
- Tools that allow users to search for other users according to the profile information they choose to make available to other users.

The first assessment was carried out in 2009 with results published in February 2010[4]. The core purpose of this second evaluation (consisting of Phase A and Phase B) is to assess how satisfactorily the signatories of the *Safer Social Networking Principles for the EU* have implemented the commitments expressed in their self-declaration reports[5] in the services they run. This analysis was made by assessing the individual self-declaration reports against the principles and, then, comparing the results of this analysis against the testing on the corresponding SNS Services.

As opposed to the first evaluation carried out in 2009, this second assessment consists of 2 Phases. In Phase A 14 "typical" Social Networking Sites were assessed in the period December 2010 - January 2011[6]. In Phase B,

[4] The results from the 1st assessment of the Safer Social Networking Principles for the EU can be consulted in the following link: http://ec.europa.eu/information_society/activities/social_networking/docs/final_report/first_part.pdf

[5] All these reports are public and can be downloaded from the European Commission's website: http://ec.europa.eu/information_society/activities/social_networking/eu_action/selfreg/index_en.htm#self_decl (link valid as of August 2009).

[6] The results from Phase A of the 2nd assessment of the Safer Social Networking Principles for the EU can be consulted in the following link:

other SNS platforms were evaluated including photo and video sharing platforms, virtual worlds, blogging and gaming platforms. These services were tested in the period March-June 2011.

Table 1 lists the 9 services included in this phase of the Second Assessment of the Safer Social Networking Principles for the EU. It includes the date of their accession to the Principles and the date of submission of the most updated version of their self-declarations available at the time of the assessment. Please see Annex 4 of this report for more detailed information on the signatories and the relevant SNS services they offer.

Signatories	**Date of accession to the Principles**	**Date of submission of the updated self-declarations**
Dailymotion	10 February 2009	10 November 2010
Habbo Hotel (Sulake)	10 February 2009	05 November 2010
Stardoll	27 August 2010	27 August 2010
Skyrock	10 February 2009	*05 November 2010*
Windows Live (Microsoft Europe)	10 February 2009	05 November 2010
Xbox Live (Microsoft Europe)	10 February 2009	05 November 2010
Flickr	10 February 2009	28 November 2010, with minor corrections made on 15 July 2011
Yahoo! Pulse	10 February 2009	28 November 2010
YouTube (Google)	10 February 2009	05 November 2010

Table 1. Signatories Participating in Phase B of the 2nd Assessment of the Safer Social Networking Principles for the EU

The first part of this report provides an overview of the implementation of the Principles by the 9 services analysed as a whole, by analyzing the main findings related to each Principle across the different services including specific examples of measures implemented. The second part of this report consists of individual reports summarising the main results of the tests carried out in each of the 9 SNS tested in this Phase.

METHODOLOGY

The 2nd Assessment of the Safer Social Networking Principles aims at determining how well the Principles each SNS committed itself to implement have been put into operation on their corresponding websites. As previously mentioned, the methodology of this 2nd assessment varies slightly in relation to 1st evaluation carried out in 2009 in that instead of testing all the SNS at once, two Phases have been foreseen. In Phase A, 14 typical SNS were tested while in Phase B (results summarized in this report) different platforms were assessed including video-sharing platforms, photo-sharing platforms, virtual worlds, gaming platform and other platforms. In order to test the videogaming platform included in this Phase, both the console and the associated website were assessed as a whole. Results of both tests were integrated into one report.

The methodology employed in Phase B consists of three main parts, namely:

(1) an analysis of the self-declarations submitted by the signatories;
(2) the testing (from a user perspective) of the websites run by the signatories, and
(3) the assessment of how satisfactory each SNS has been in the implementation of their individual commitments in relation to the Principles (as expressed in their individual self-declarations) on the services they run.

---

http://ec.europa.eu/information_society/activities/social_networking/eu_action/implementation_princip_2011/index_en.htm

In the second assessment all the individual self-declarations were analysed solely by the coordinator so as to ensure a maximum degree of objectivity during the assessment. During this second assessment national researchers carried out the tests on each SNS. These tests only evaluated Principles 1,2,3,4 and 6. Because of ethical reasons Principles 5 and 7 were not tested on the websites.

As opposed to Phase A where a standard questionnaire was used for testing all the websites, in Phase B an adapted version of the questionnaire was designed for each type of platform being tested. This responded to the need of taking into account the specific nature and diversity of each of the platforms being evaluated in this Phase. In total 5 different versions of the testing scenario were developed: for video-sharing platforms, photo-sharing platforms, virtual worlds, gaming platforms and for other platforms. The analysis of the testing scenarios and the reporting focused on those aspects each service had committed itself to implement in their individual self-declaration.

The coordinator analysed the results of the national tests. The way each principle had been implemented was assessed according to what was stated in each individual self-declaration. Three categories of assessment were established, namely, *very satisfactory, rather satisfactory and unsatisfactory*. As observed later on in the results section, it is possible that in a few cases some services, whose self-declaration was assessed as "unsatisfactory" or "rather satisfactory", may have been evaluated as "(very) satisfactory" in the implementation on their website. This is the case of those providers whose self-declaration does not provide thorough information on each and every measure they have actually implemented on the websites they run.

STEP 1: ANALYSIS OF THE SELF-DECLARATIONS

The first step of this assessment consisted in determining if each individual self-declaration was in line with the Safer Social Networking Principles. Each signatory's self-declaration was assessed by means of a questionnaire consisting of seven sections, each of which was devoted to evaluating one of the seven Safer Social Networking Principles[7]. Each section contained a series of questions that aimed to assess how well each SNS had addressed the most important facets of each Principle in their self- declarations.

At the end of each section of the questionnaire, the coordinator filled in a table summarizing how well each Principle was implemented in the corresponding self-declaration ("*very satisfactorily", "rather satisfactorily"* or "*unsatisfactorily")* as well as areas for further improvement.

Finally, the analysis of the self-declarations also included a section where the coordinator referred to any relevant additional information/features/functionalities that the company may have included in their individual self-declaration but which were not included in the current version of the questionnaire.

STEP 2: TESTING THE SNS WEBSITES

Following the analysis of the self-declarations, in Phase B each SNS was tested by one national researcher in the (main) local language of the signatories' website with the exception of Habbo Hotel which was tested in both Finnish and English. In the latter case, both testing results were merged into one report summarising the main findings of both tests carried out on the website.

In order to create the right assessment instrument the testing questionnaire employed in the first assessment was reviewed and improved[8]. In concrete terms, 5 different testing scenarios were developed, one for each of

[7] For a full description of each of the Safer Social Networking Principles, please consult: http://ec.europa.eu/information_society/activities/social_networking/docs/sn_principles.pdf

[8] The questionnaire employed during the 1st assessment of the Safer Social Networking Principles for the EU can be consulted in the following link (Annex 3): http://ec.europa.eu/information_society/activities/social_networking/docs/final_report/first_part.pdf

the platforms being tested. Each of these questionnaires included common as well as specific items specially adapted to the platform being assessed. Besides, several statements were reworded or refined. To the extent possible, factual questions were used to operationalize subjective terms such as *"accessible*" or "*easy-to-understand*".

Each of the five final testing instruments developed for Phase B comprise 5 sections, one for each of the principles assessed (Principles 1,2,3,4 and 6). Each section contains a set of questions (indicators) that evaluate the extent to which each SNS has addressed the Principles (as stated in their individual self-declarations) on their websites. Alternative ways of implementing the Principles were also taken into consideration. Indeed, national researchers were encouraged to refer to any alternative ways any particular Principle had been implemented on the website and which had not been covered by the testing scenario.

The methodology employed for this second assessment was based on the "mystery shopper" technique, where researchers had to set up profiles of minors and adults to carry out the tests. Together with the testing scenario all national researchers were provided with a set of "personas" which served to create the necessary profiles to carry out the tests. A persona is an archetypical representation of real or potential users. The persona, thus, represents patterns of users' behaviour, goals and motives, compiled in a fictional description of a single individual. It also contains made-up personal details in order to make the persona more "tangible and alive" for the testing team. In our testing scenario, the personas fulfilled two main roles:

(1) They provided relevant information (including a set of pictures) that was used to create realistic user profiles for the tests. By doing this we could also ensure that all the tests were carried out under similar conditions.

(2) Because the persona model resembles classical user profiles, they can help to develop empathy with the target users. In this specific case, they were used to help adult national researchers think from the perspective of a child and/or an adolescent user.

The methodology allowed making a thorough and comprehensive assessment of each SNS individually. At the same time, as several questions were common to all testing scenarios, it was possible to merge much of the data obtained into a common data set and conduct an overall analysis of the most relevant issues per Principle across all the SNSs evaluated.

Undoubtedly, this methodological approach also has its limitations. Arguably, the main one being that no minors were included in the testing process and thus, all the results presented in this report are solely based on the expert assessment carried out by adults. Even though a big effort has been made to ensure the consistency and validity of the methodology employed for this assessment, it is important to keep in mind that some aspects of the principles will necessarily be better evaluated with real children actually interacting with each of the websites. However, the use of Personas helped to counteract to some extent this problem by supporting researchers think and make decisions which may be closer to the ones younger users would have actually made. Another limitation of this approach lies with the fact that all the tests (with the exception of the website that was tested in 2 different languages) were carried out by only one national researcher. Research in other fields (such as usability) has demonstrated that the results from expert evaluations can be maximised when expert evaluations are carried out by a group of experts (at least 2) rather than by one individual (Nielsen, 1990). This second shortcoming was overcome by the close collaboration between national researchers and the coordinator. The latter supported national researchers throughout the whole testing and analysis period and was consulted by national researchers whenever problems related to the analysis or interpretation of the data arose. Finally, it is important to mention that the testing was carried out in the period March - July 2011 and therefore the results contained in this report refer to that period only.

GENERAL FINDINGS

The first step of this assessment consisted in determining if each individual self-declaration was in line with the Safer Social Networking Principles. This analysis revealed that self-declarations of 2 services were *very satisfactory* in all the Principles assessed (See Fig. 1). 1 service was assessed as *very satisfactory* in 6 Principles, 3 services were *very satisfactory* in 5 principles and 3 services were *very satisfactory* in 4 Principles.

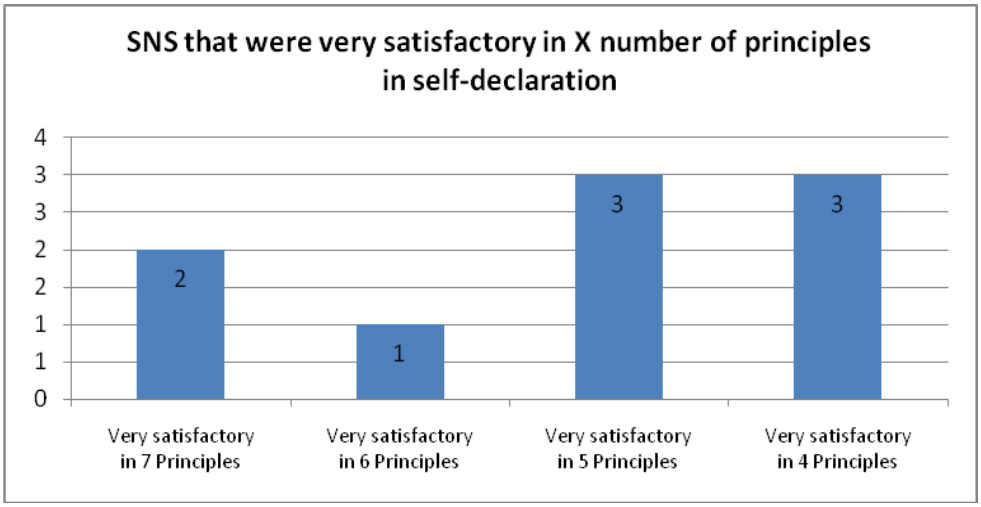

Fig. 1 Services assessed as very satisfactory in X number of Principles in self-declaration

When looking at how satisfactory the implementation of the self-declaration statements on the respective services was, Figure 2 shows that two service providers were assessed as *very satisfactory* in relation to the implementation of their commitments as expressed in their individual self-declaration on the 5 Principles tested on the website. 5 services were assessed as *very satisfactory* in 4 of the 5 Principles tested, 1 service was assessed as *very satisfactory* on 3 Principles and 1 service was assessed as very satisfactory in 2 Principles.

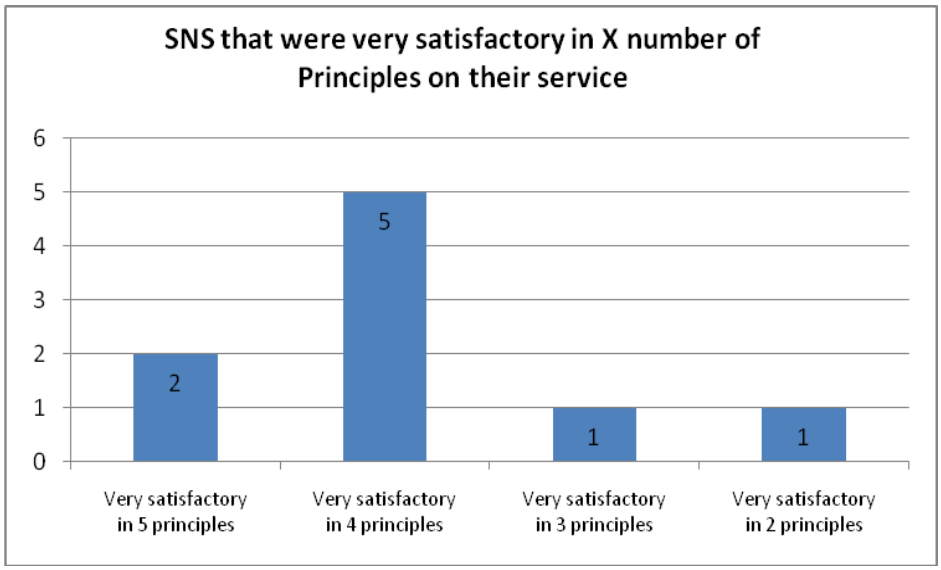

Fig. 2 SNS assessed as very satisfactory in X number of Principles on their service

The analysis of the self-declarations of the 9 Social Networking Sites evaluated shows that Principle 5 "Assess the means for reviewing illegal or prohibited content/conduct" and Principle 7 "Assess the means for reviewing illegal or prohibited content/conduct" were the best assessed in terms of their self-declarations with all the services assessed *very satisfactorily* (See Fig. 3). Principles 1 and 2 were also very well assessed with 8 out of 9 services being assessed as *very satisfactory.* Because of ethical reasons Principles 5 and 7 were not tested on the websites.

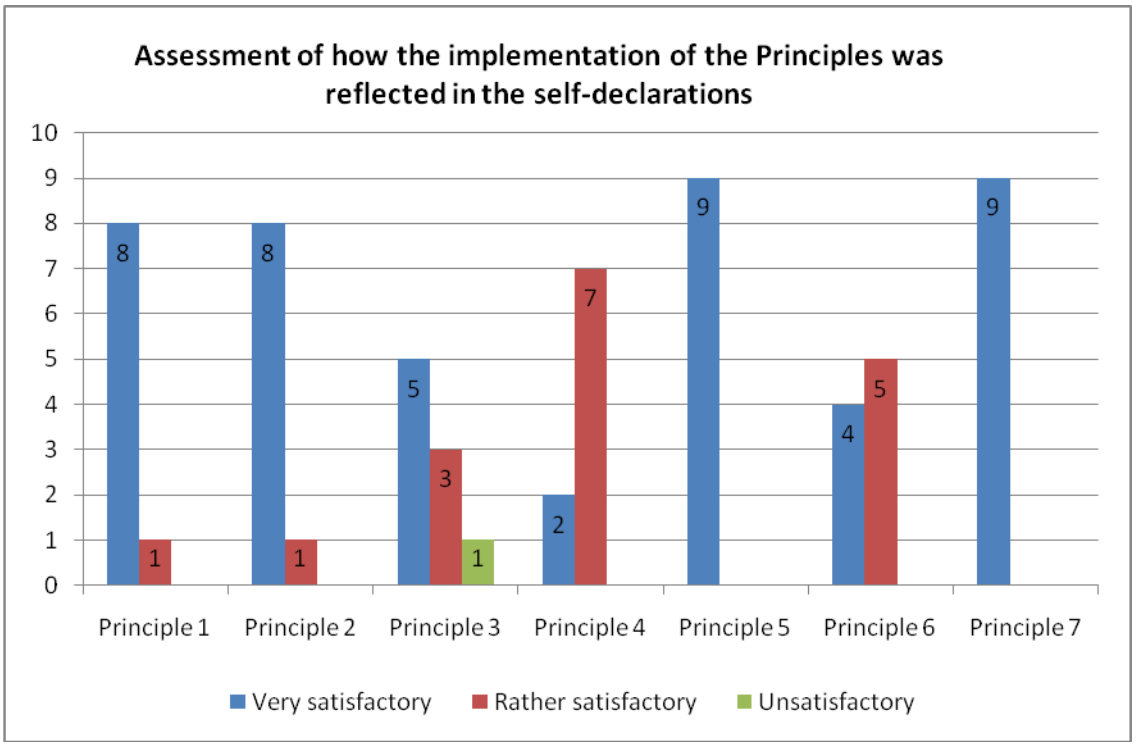

Fig. 3 Assessment of how the implementation of the Principles was reflected in the self-declarations

MAIN FINDINGS IN RELATION TO THE IMPLEMENTATION OF THE INDIVIDUAL SELF-DECLARATIONS ON THEIR RESPECTIVE WEBSITES

As part of this assessment, all Principles, except 5 and 7, were tested through a wide range of tasks that national researchers had to perform on the websites of the 9 SNS being evaluated in this Phase. In terms of the implementation of the individual self-declarations on the services tested, Principles 1 "Raise awareness" and Principle 4 "Easy mechanisms for reporting violations" were the best assessed. Principle 3 "Empower users" was the least well evaluated on the website, although it is worthwhile mentioning that no service was assessed as *"unsatisfactory*".

The following sections summarise the main findings by Principle across the 9 services tested.

Principle 1: Raise awareness of safety education messages and acceptable use policies to users, parents, teachers and carers in a prominent, clear and age-appropriate manner

Principle 1 states that Social Networks should *"Raise awareness of safety and education messages and acceptable use policies to users, parents, teachers and careers in prominent, clear and age-appropriate manner"*. The principle is operationalized into five specific recommendations ("Safer Social Networking Principles for the EU," 2009: 6):

- Providers should create clear, targeted guidance and educational materials designed to give children and young people the tools, knowledge and skills to navigate their services safely.
- These messages should be presented in a prominent, accessible, easy-to-understand and practical format.
- Service providers should provide clear information about what constitutes inappropriate behaviour. This information should be easily accessible and include information about the consequences of breaching these terms. Providers should explore other ways to communicate this information outside the terms of Service.
- Providers should offer parents targeted links, educational materials and other technical controls as appropriate with the aim of fostering dialogue, trust and involvement between parents and children about responsible and safer internet use.
- SNS providers should ensure that such materials also empower teachers to help children use SNSs safely and responsibly.

Overall, the assessment of Principle 1 is very positive on the website and consistent with the providers` commitments in their self-declarations. All the SNS assessed were evaluated as *very satisfactory* on their services

All the services fulfilled their commitment expressed in their self-declaration including in their services safety information, guidance and/or educational materials specifically targeted at children, their parents or guardians and teachers. In all these services this information was easy clear for minors to understand. In 7 services this information was also easy to find. These services also provided general Terms of use and/or adapted child-friendly versions thereof (e.g. via Community Guidelines, User Agreements, House Rules, etc.).

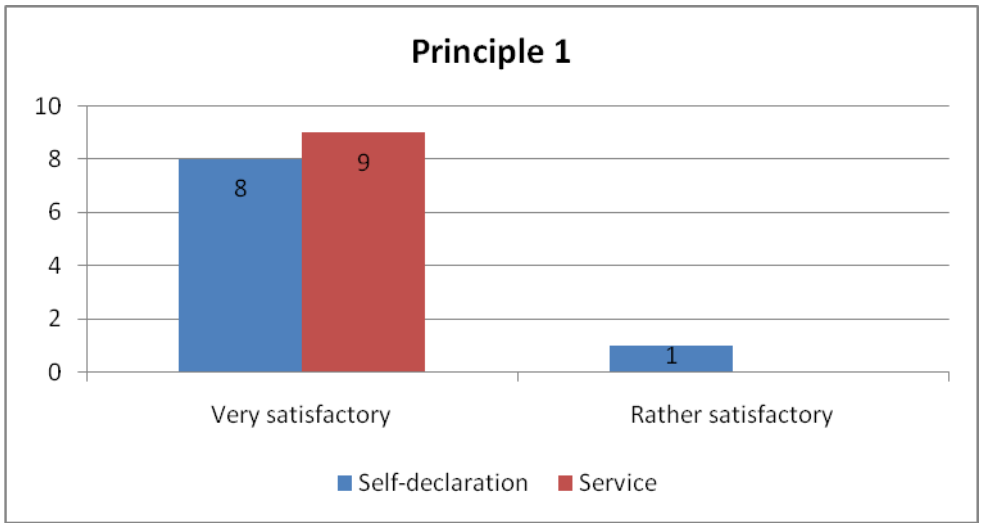

Fig. 4 Assessment of Principle 1 in self-declaration vs. service

WHAT TYPE OF SAFETY INFORMATION/MATERIALS ARE AVAILABLE?

All the services, but one, provide safety information, guidance and/or educational materials on their websites specifically targeted at children (see Fig. 5). Only in one of the services the information provided was not specifically targeted at minors, but rather at their parents or guardians. Nevertheless, the service in question provides plenty of general safety information and tips throughout the site which could also be easily understood by minors.

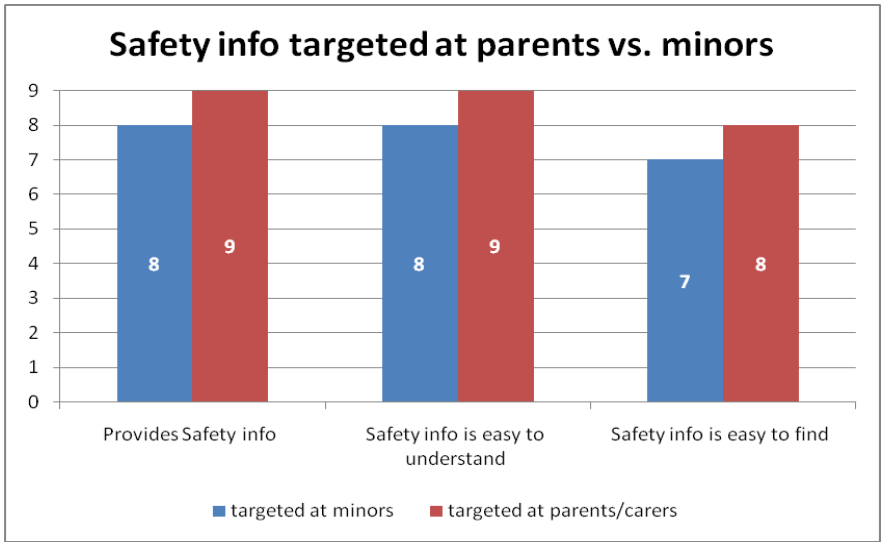

Fig. 5 Availability of safety information for minors and their parents or guardians

Regarding the type of safety information provided, all the services include general safety information. 8 services include specific information about pornography or sexual content, bullying, hate speech and risks associated to divulging personal information on their services. 6 services provide information on the risks associated to posting sexually provocative pictures and 5 include information on violence and adults with sexual interest on children. Information on other topics such as self-harm (suicide, anorexia, bulimia, etc.) is less common (See Fig. 6).

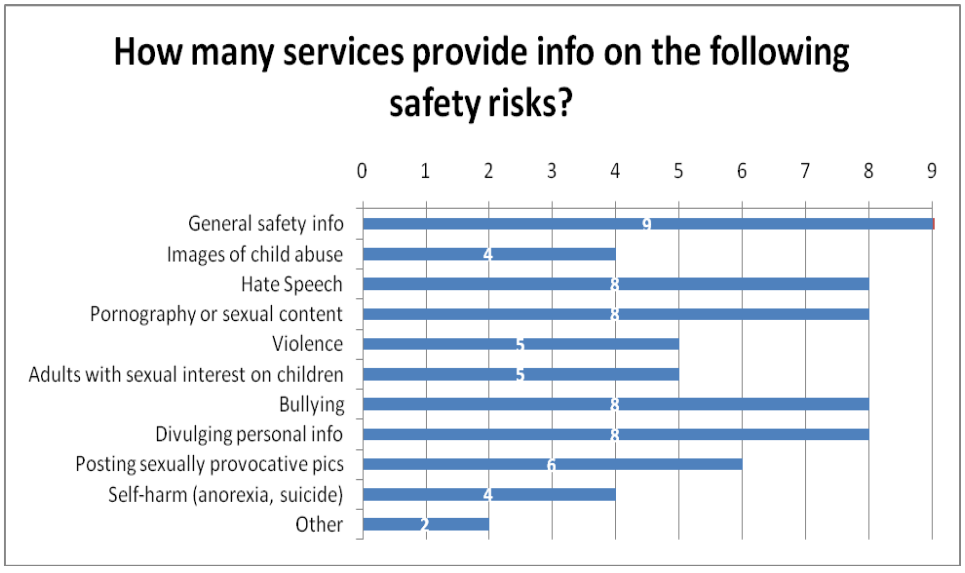

Fig. 6 Number of services that provide information on the following safety risks

Regarding the format of the safety information available for minors, in the majority of the cases this information was presented in the form of written texts, external links or referrals to (educational) organizations active in child safety and concrete examples related to safety. In 6 services, this information is also provided via audio-visual fragments and only 1 service presented this information via games (See Fig. 7).

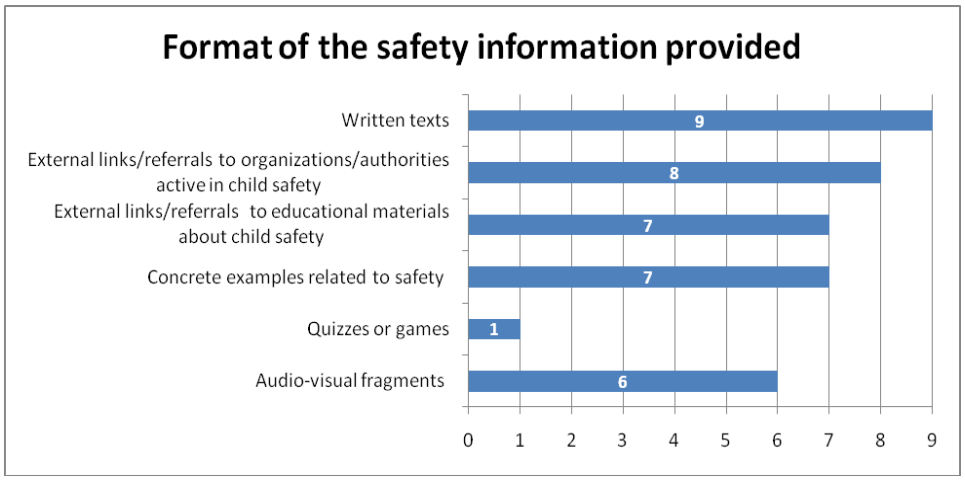

Fig. 7 Format of the safety information provided by the services tested

IS THE SAFETY INFORMATION PROVIDED BY THE SNS EASY TO FIND AND TO UNDERSTAND?

In all but one service (which did not provide safety information specially targeted at children, but general safety information targeted at parents and carers), **was the information related to safety clear and age-appropriate and easy to understand for minors**. **Safety information for minors was easy to find in 7 of these services**. Excluding the new comers, findings regarding this principle are similar to the ones of last year's assessment.

The safety information and tips for teachers and parents were clear and easy-to-understand in all the services assessed and they were easy to find in 8 of the 9 services that provide it (See Fig. 5 above). Some difficulties were encountered in locating safety information in those services that offered (additional) safety information via corporate websites rather than in the local websites/services themselves. Once found this information proved to be comprehensive and relevant, though.

IS IT CLEAR TO USERS WHAT CONSTITUTES INAPPROPRIATE BEHAVIOUR ON THE SNS?

All the SNS provide general Terms of Use or Service as well as an adapted, shorter and more child-friendly version of the Terms. This shorter version is usually presented in the form of Community guidelines, User Agreements or House rules. The general Terms of Use provided by the services were in all cases, but one, assessed as difficult to understand by minors. This is mainly because they are usually long, legalistic texts that contain technical terms which may be difficult for younger audiences to understand. However, **all the services assessed provided an adapted version of the Terms which was easy-to-understand and easy to find.** Only one service provided both an easy-to-understand version of the terms of Use and an additional, adapted, child-friendly version of such Terms. Both the general version of the Terms of Use as well as the shorter child-friendly versions thereof include explicit information on what constitutes inappropriate or forbidden behaviour on the service (e.g. uploading illegal or inappropriate materials such as pornography, exploiting minors in any way, harassing other users, etc.) and the consequences thereof (e.g. being reprimanded, having one`s account suspended, temporarily cancelled or deleted, or even being reported to the corresponding legal authorities depending on the severity of the offence committed).

Regarding the format of the Terms of Use, User Agreements or Community guidelines available for minors, in the majority of the cases this information was presented in the form of written texts, external links or referrals to relevant organizations and/or concrete examples that illustrated the conditions of use of the service. Only in one service, this information was also provided via audio-visual fragments.

EXAMPLES OF BEST PRACTICE FOR PRINCIPLE 1:

Examples of Best Practice regarding the implementation of Principle 1 on the services tested are Dailymotion, Habbo Hotel, Flickr, Yahoo! Pulse, Skyrock, Stardoll and YouTube. All these services provide targeted safety information for children as well as for parents and carers. The information they provide was easy to find and to understand. Besides, all these services provide a child-friendly adapted version of the Terms of Use governing the site.

Principle 2: Work towards ensuring that services are age-appropriate for the intended audience

Principle 2 states that Social Networks should *"work towards ensuring that services are age-appropriate for the intended audience".* In order to assess the implementation of such services, a differentiation has been made between 1) restrictions meant to ensure that those below the intended minimum age of the service cannot register (sign-up restrictions), and 2) restrictions aimed at ensuring age appropriate services when the user is already registered and a member of a social networking site.

According to the self-declaration, Principle 2 was evaluated as *very satisfactory* on 8 services tested and as *rather satisfactory* in one service. On the websites, Principle 2 was assessed as *very satisfactory* in 7 services and as *rather satisfactory* in 2 services (See Fig. 8). The best evaluated services where those where their commitment to deny access to underage users was in place, where effective mechanisms to prevent re-registration as expressed in their self-declarations (e.g. employing cookies, demanding parental permission to register, etc.) were in place and where minors were not confronted with any type of inappropriate content including advertising (in the case of those services that referred to specific measures related to advertising in their self-declarations).

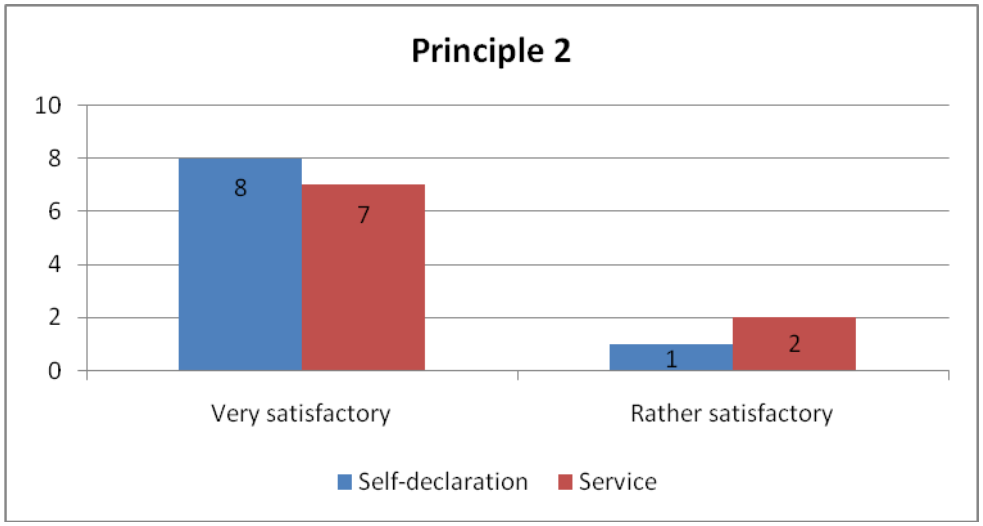

Fig. 8 Assessment of Principle 2 in self-declaration vs. service

ARE THE SERVICES AGE-RESTRICTED?

5 of the 9 SNS have set up a minimum age requirement in order for users to be able to sign up to their services. In 1 of these services the minimum age requirement is 12 and in 3 services it is 13 years old. In the service tested in two language versions, the minimum registration age was 10 in the Finnish version and 13 in the UK version of the site.. Within the services that are age-restricted the most common age verification mechanism is self-declaration of age. In 4 services no minimum age registration applies. However, in one of these services, parental consent is needed to approve a Community Membership of the site for all children

under the age of 13. In another non age-restricted service, accounts for under 18 year olds need to be set up by an adult. This is verified by entering valid credit card details. In the latter case, parents are also required to be responsible for child or teen accounts and to authorise any amendments to default settings.

WHAT HAPPENS IF AN UNDER-AGE USER TRIES TO SIGN UP?

In order to test this, researchers attempted to create a profile of a 9-year old in the 5 services where a minimum age registration applied. When attempting to register, new users were demanded to provide some basic personal information including their name or username and date of birth. Researchers were instructed to provide the child`s "real" date of birth at registration, i.e. 9 years old. The results of this test showed that **all the services intended to be age-restricted effectively prevent sign-up by underage users on their sites.** These services state in their self-declaration that they employ mechanisms to identify underage users from their services including user reports and searching underage users manually. They also indicate that when an underage user is identified their accounts are deleted from the service.

IS IT POSSIBLE FOR MINORS TO RE-REGISTER ON THE SITE ONCE THEY HAVE BEEN DENIED ACCESS FOR BEING UNDER AGE?

If signing up as a 9-year-old child had failed researchers were instructed to attempt to re-register on the site indicating that their age was just above the minimum age required by the SNS. Researchers were instructed not to erase any cookies nor modify any of their previous settings. Testing revealed that the **5 services intended to be age-restricted prevent (initial) sign-up by underage users on their sites, although in 2 of these services users could eventually re-register on the site** by simply changing their initial age to one just above the minimum age required by the SNS. In one service cookies had been installed to avoid re-registration, so sign-up was possible only after removing these cookies. This was achieved, for instance, by closing the browser and opening it again and then changing the date of birth to one just above the minimum age required by the service. In the other 2 age-restricted services, it was not possible to simply change the minor's age to be able to register. However, it must be noted that as no 100% reliable age-verification mechanism exists up to date, even in these latter cases where age-restriction mechanisms proved to be effective, it was possible for underage users to register on the site by creating a completely new account of a user with an age above the minimum required by the service.

ARE MINORS CONFRONTED WITH ANY TYPE OF INAPPROPRIATE CONTENT AND/OR SERVICES?

Practically none of the services tested included "adult" sections or services, except one where content labelled as "explicit" (including erotic or sexually explicit videos) was inaccessible to minors by default. In all the other services tested all services and sections were appropriate for all audiences.

Regarding the appropriateness of content found on the services, this depends largely on the adequate tagging or labelling of the material uploaded by users themselves. This is specially the case of photo-sharing and video-sharing platforms due to the large amount of content uploaded every minute. Services, therefore, have implemented community-driven mechanisms such as tagging, labelling and flagging of content to support the classification and age-restriction of materials uploaded to the sites. The photo-sharing and video-sharing platforms tested in this Phase provided user-friendly mechanisms to label and categorize content so that "safe" or "child-appropriate" videos and pictures could be easily distinguished from adult content. Common categories to age-restrict content included labels such as "moderate", "restricted", "sexy", "18+", etc. The classification options varied from service to service but they all included an option to restrict content which may be considered as not appropriate for minors. During testing **labelling or tagging mechanisms (associated to internal filtering processes) proved to be effective, although not infallible because in a few cases it was possible to identify content that could be considered as inappropriate for younger children or adolescents**, including drug paraphernalia, self-harm triggering and highly erotic content (clearly not meant for artistic or educational purposes).

IS ADVERTISING AGE-APPROPRIATE?

Although advertising is not included in the Safer Social Networking Principles for the EU, 3 of the 9 services tested explicitly indicate in their self-declarations that advertising on the services they run is always age-appropriate. In these 3 services, no advertising that could be considered as inappropriate for minors was found.

WHAT MECHANISMS ARE IN PLACE TO DETECT AND RESTRICT ACCESS TO INAPPROPRIATE CONTENT?

5 services mentioned in their self-declarations that text or content filters were employed to limit the exposure of minors to inappropriate language or content on their services. In 3 services users were prevented from uploading pictures or videos of real people and in one service all images were pre-moderated before appearing on the site. Other common available "opt-in" functionality especially in the case of photo and video-sharing platforms was the use of "safe" or "child-appropriate" mode meant to restrict access to adult content. In the services that offered this type of content-restricting mechanism, a series of image and video searches were conducted (with safe mode enabled) using terms which could expose minors to inappropriate content (e.g. sex, violence, nudity, drugs, gore, porn). These searches generally did not lead to problematic content. There were some nude images which would be classified as artistic, as well as some photos of drug paraphernalia (e.g. things for smoking marihuana), but no pictures of drug taking as such. In a few cases, though, erotic and triggering self-harm content was identified.

Services encouraged users to report any illegal content found on the service as well as wrongly labelled content (e.g. when a video or picture displays inappropriate content for minors, but is classified as "appropriate" for all ages). During testing, one image which contained pornographic material within the photo on a computer screen was identified in one of the services. This image was reported to the customer care team of the service where the picture was found and it was, indeed, reclassified so that it was no longer available for minors.

ARE "SAFE" MODE AND SETTINGS EFFECTIVE IN PREVENTING MINORS FROM ACCESSING INAPPROPRIATE CONTENT?

5 services offered "safe" mode settings; however, these settings were not always enabled as a default on the accounts of minors. With these safety settings enabled, age-restricted content should not show up in video or photo searches or related user-generated content. **In a few cases, even though safe mode was enabled by default, it could be modified or even be disabled by minors**. For instance, in one service it was possible for minors to change their safe default setting to enable access to content rated as "moderate" (although access to content rated as "restricted" was not allowed on this specific service). Setting the safe mode to "moderate" provided access to some "inappropriate" content that was clearly neither artistic nor educational, including an image of a man with his genitals on display. In another service the safe mode could be set to "off" if the minor clicked on a link indicating that he was older than 18 years old. This provided access to explicit content mainly of a sexual nature. A few services offer additional parental controls that allow, for instance to lock safety settings to ensure that they can't be switched off by minors using the computers where the parental controls have been installed.

ARE MECHANISMS TO MODERATE INTERACTION ON THE SERVICE EFFECTIVE?

In two of the services tested, online communication was real-time moderated both through the use of text filtering and by staff. In an experiment set up to test the efficiency of the moderation mechanisms, accounts were set up for an adult (male, 25 years) and a minor (female, 15). The adult account was set up specifically to seek contact information from a minor. In both services tested, it was possible for the adult account holder to make contact with the minor via hosted chatrooms. In one of the services, divulging personal information such name of school attended or phone number was blocked. Text filtering was also successfully applied when racist comments or inappropriate language was posted in comments on some of the sample accounts set up. In the other service tested the adult obtained some information (place of residence, school and address), but phone numbers or e-mail addresses were replaced by asterisks on the screen. When the minor tried to give a phone

number for the second time, the moderator issued a warning on the adult's screen explaining that enquiring or giving phone numbers was forbidden according to the rules of the site.

ARE PARENTAL CONTROL MECHANISMS EFFECTIVE?

5 of the services tested employ some form of parental control on their website. These mechanisms vary greatly from service to service. While some services provide sophisticated parental control tools that allow parents or guardians to effectively monitors their children's use of the service, other services employ basic control mechanisms such as informing parents that their child has opened an account on the service or asking for parental consent during registration.

Only one of the services tested provided a sophisticated, but still user-friendly downloadable software application. The software needs to be installed by a parent on each computer used by a child in order for parental controls to take effect. With the use of parental controls, it is possible, for instance, to limit a minor's internet access to just child-safe websites; to require approval for online friends; to specify who is able to contact the child such as by email or instant messaging; and to block access to other designated websites (e.g. other social networking sites). For the purposes of testing, this parental control software was set up for a child account. A range of settings were tested including "Child Friendly", which blocks adult sites and online communication; "General Interest" which allows a wider range of web access but blocks social networking; and "Online Communication" which allows social networking but also blocks adult sites. One of the tests conducted consisted in limiting access to child-friendly websites suitable for a nine year old and blocking access to other sites that might not be deemed age-appropriate. **All of these controls were found easy to install and operate; they proved to be effective in providing a high degree of parental control over all aspects of online search, web access and online communication.**

EXAMPLES OF BEST PRACTICE FOR PRINCIPLE 2:

Flickr, Habbo Hotel, Yahoo! Pulse, Skyrock and YouTube prevent (initial) sign-up of minors on their services. In Flickr and Yahoo! Pulse minors could not re-register on the site even if they changed their registration age to one above the minimum required by the service.

In Stardoll even though no minimum age requirement applies, parental consent is required to approve a community membership of the site for all children under the age of 13. In addition, when registering on the site, those under 13 automatically get a special category of membership called "Kid Safe" that blocks communication with other users of the site. In Xbox Live accounts for under 18 year olds need to be set up by an adult and are verified by entering valid credit card details. Parents are also required to be responsible for child or teen accounts and to authorise any amendments to default settings.

Windows Live provides free parental controls that were found easy to install and operate; they proved to be effective in providing a high degree of parental control and monitoring over all aspects of children's online search, web access and online communication.

Principle 3: Empower users through tools and technology

The third principle "*Empower users through tools and technology"* refers to the tools and technologies employed to assist children and young users in managing their experience of the service. According to the "Safer Social Networking Principles for the EU" (2009: 7-8) such tools include:

- Taking steps to ensure that private profiles of users registered as under the age of 18 are not searchable
- Set private profiles for users below 18 to private by default
- Make private profiles viewable only to "friends"/people on the user's contact list

- Give users control over who can access their full profile
- Give users control over who can post comments and content on their profile and the possibility to delete messages and other content
- Give users the option to pre-moderate comments from other users before they are published on their profile
- Provide easy-to-use tools for reporting inappropriate contact or content from other users[9]
- Educate parents about available tools.

Principle 3 was assessed as *very satisfactory* in 5 individual self-declarations, as *rather satisfactory* in 3 and as *unsatisfactory* in 1. On the websites, Principle 3 was assessed as *very satisfactory* in 5 services and as *rather satisfactory* in 4 services (See fig.9).

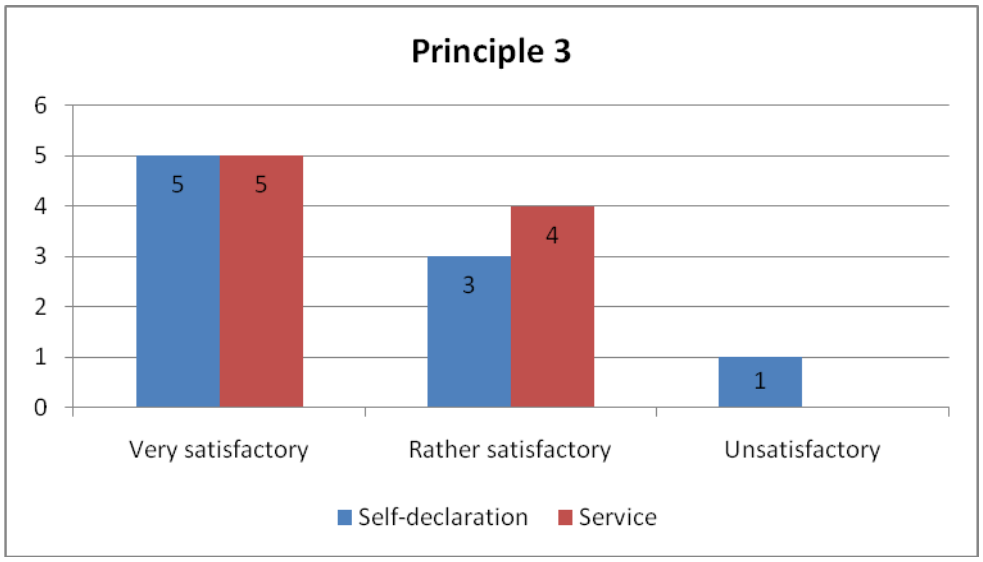

Fig. 9 Assessment of Principle 3 in self-declaration vs. service

WHAT TYPE OF INFORMATION IS VISIBLE TO USERS BEYOND THE MINOR`S APPROVED LIST OF CONTACTS?

According to the analysis of the self-declarations, 3 services explicitly state that the default settings for users under 18 years old are set to friends only. Of these three services, only one does not display personal, identifiable information from the minor to (non) users beyond the minor`s approved contact list by default. In one of these services, even though limited personal information was displayed, it was still accessible to friends of friends by default. In the other 6 self-declarations analysed, it is not completely clear what type of information is visible to other users by default. Testing on the services demonstrated that in only one of these 6 services, profiles of minors do not display personal, identifiable information from the minor to (non) users beyond the minor`s friends by default. Thus, in sum, **only 2 services make minors' personal and identifiable information visible by default only to their approved list of contacts.** This means that even if minors add extra information about themselves (not required during registration) to their profiles this information is not displayed to users beyond the minor's contacts by default.

Regarding the information contained in user profiles, several services state in their self-declarations that very little personal information from minors is required at registration. However, some of these self-declarations also state that if users wish so, they may add more personal information to their profiles. It is not clear from the self-declarations, though, if this optional information is (automatically) mapped into the user's profile and,

[9] For elaboration on the implementation of this point, please refer to Principle 4 and Principle 5.

thus, made visible to other users (beyond the minor's approved contact list) by default. Testing, however, demonstrated that a considerable amount of personal information -including information added by users after registration- was displayed to users beyond the minor`s approved contacts list by default. In some cases, the (full) name of the minor, gender, age, the school they attend, the minor's location (although not their address), profile picture, pictures and videos uploaded by the minor or their contacts, comments posted on the minor`s profiles, the minor`s contact list or online status, their interests and/or hobbies, etc. were displayed to either "friends of friends" and/or non-friends. It is important to stress, however, that this information did not always lead to the identification of minors and, therefore, does not necessarily constitute an imminent safety risk. Furthermore, platforms tested in Phase B are not "typical" social networking sites. This means that in several services user profiles are not the main point of entry or of interest for a user. This is specially the case of photo-sharing and video-sharing platforms where users would typically visit a photostream or a video channel, but not necessarily a user's profile.

Compared to last years` results, no improvement was observed regarding the implementation of this measure.

WHO CAN GET IN TOUCH WITH MINORS?

In this report we distinguish among three ways of getting in touch with minors, namely via friend requests, via public messages (e.g. comments posted on the minor's public profile) and via private messages. According to the analysis of the self-declarations, 3 services explicitly state that the profiles of minors under the age of 18 are defaulted to connections only. One service states that members under the age of 13 are suggested to sign up for a Kid Safe account that does not allow them to communicate with other members of the community, but no specific measures are mentioned for users older than 13. The other 5 self-declarations analysed do not explicitly state if minors can (or cannot) be contacted by users beyond their approved list of contacts. Nevertheless, they do state that users (including minors) can control how others interact with them (or with their content), for instance, by allowing them to block specific users, to restrict interaction, to reject friends' requests, etc.

Of the 3 services that self-declared that the profiles of minors under the age of 18 are defaulted to connections only, in only 1 service it was impossible for non-friend users to get in touch with minors via private or public messages by default. In the other two services the possibilities for contacting minors were very limited, but still in one of these services it was possible for other non-friend users to make comments in the publicly accessible updates of minors and in the other service minors who used the forum feature could join in conversations, start new threads and communicate with other non-friends members via private messages. In one of the services that didn't self-declare it minors could not be contacted by anybody beyond the minor`s approved contact list via public nor private messages by default. Testing further revealed that, by default, **by default, in all the services tested minors could be contacted via friend requests, in 7 services they could be contacted via public messages and in 3 services they could be contacted via private messages by users beyond their approved list of contacts.** In all the services tested, though, privacy settings allow minors to control, at least to some degree, who can get in touch with them, for instance, by blocking other users or by setting specific communication restrictions (e.g. only friends can post comments on the minor's profile). It must be noted too that in one of these services, only 17 year old users could be contacted by non-friends by default (including non-registered users) via private messages or by posting comments on the minor's public blog. However, users younger than 17 (12 and 16 year olds in the case of our tests) could only be contacted by friends. In another service, where minors could, indeed, be contacted by non-friends via public messages, children under 13 years old and who have signed up for Kid Safe account, are not allowed to access any of the communication functions of the site. In one of the 3 sites where users could only be contacted via friend requests, these can only be accepted with parental consent.

CAN PROFILES OF MINORS BE FOUND BY USERS BEYOND THE MINOR`S APPROVED LIST OF CONTACTS?

By default, in only one of the 9 services tested profiles of minors could be found by name searches via internal and external search engines (e.g. Google, Bing or Yahoo!). This was tested by typing the name of the minor between inverted commas (e.g. "John Smith") in several search engines. Nevertheless, this did not

guarantee that the profile of minors could not be found via other mechanisms in the other services tested. For instance, **in 6 of the services tested the profiles of minors could be directly accessed via the profiles of the minor's friends**. In these cases, friends of friends have, by default, access to their friend's contact lists and can, thus, get access to profiles of minors who do not belong to their own lists of contacts. In this sense, profiles of minors could be considered as "unlisted", i.e. they do not appear in searches but may be viewed by anyone with a link to the profile.

User-generated content is another way for people (including sometimes non-registered users of the sites) to get access to minor's profiles. For instance, in one of the services tested, photos identified in searches provided access to the username of minors, gender, relationship status, likes and dislikes, and their photostream to registered users and non-users of the site. In other words, even though the profile cannot be *searched* via search engines, it can still be *found*. Once found, considerable information from the minor was displayed to people beyond the approved contact list of the minor by default. Besides, all registered users of the site could make comments on photos, send friend requests and even private e-mails. A similar situation was observed in the video-sharing and blogging platforms tested. Once the minor's blog, individual videos or video channels were found (e.g. via an external search engine), access to the minor's profile was granted. In all these cases, the amount of personal information displayed on the profiles depends on the amount of information provided by the minor as well as on the default settings and the additional privacy settings that the minor may have decided to set up. In one of the virtual world platforms tested, user profiles are searchable within the service, but this is only successful if the username is known (not the real name). This information could be determined by seeing another user in a chatroom and searching their username, or if they publicly posted their username somewhere else online. In this service, the profile pages of all users are set to public by default. However, the profile page only contains the username and online status in default setting what limits the amount of personal information they contain. Besides, it was forbidden to post personal information on the profile page (e.g. pictures, videos, etc.).

CAN USERS CONTROL WHAT IS POSTED TO THEIR PROFILES?

Deleting unwanted comments was possible in the 8 services that allowed users to post comments on each other's profiles. In one service this functionality was not available. By default, only in half of the services that allow posting comments friends were the only ones allowed to post them. In the other 4 services, non-friends could also post comments on the profiles of minors. In the 9 services tested users could block other users and reject friends 'requests.

EXAMPLES OF BEST PRACTICE FOR PRINCIPLE 3:

Habbo Hotel and Xbox Live are the only 2 services that make minors' personal and identifiable information visible by default only to their approved list of contacts. Even if minors add more information about themselves (not required to register to the service) this information is not displayed to users beyond the minor's contacts by default.

In Dailymotion and Windows Live minors cannot be contacted by anybody beyond the minor`s approved contact list via public nor private messages.

Principle 4: Provide easy-to-use mechanisms to report conduct or content that violates the Terms of Service

In order to successfully implement principle 4 *"Provide easy-to-use mechanisms to report conduct or content that violates the terms of service"* on their services,

- Providers should provide a mechanism for reporting inappropriate content, contact or behavior as outlined in their Terms of Service, acceptable use policy and/or community guidelines. These mechanisms should be easily accessible to users at all times and the procedure should be easily understandable and age-appropriate.

- Reports should be acknowledged and acted upon expeditiously.
- Users should be provided with the information they need to make an effective report and, where appropriate, an indication of how reports are typically handled.

According to the self-declaration only 2 services were assessed as *very satisfactory* and 7 *as rather satisfactory*. On the website, 8 services were assessed as *very satisfactory* and only 1 as *rather satisfactory* (See Fig. 10).

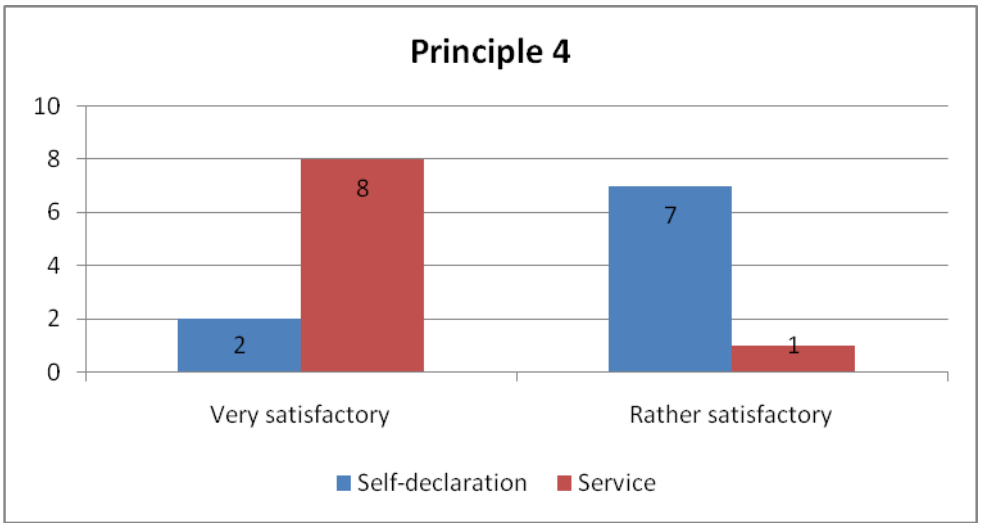

Fig. 10 Assessment of Principle 4 in self-declaration vs. service

WHAT REPORTING MECHANISMS DO SNS OFFER TO REPORT INAPPROPRIATE CONTENT OR CONDUCT ON THEIR SERVICES?

All the services tested provided one or more mechanisms to report inappropriate content or contact on their website including a general report button and/or a report button next to user-generated content (e.g. to flag inappropriate pictures, videos or comments). 7 services provide both. Report forms were available in all the services and reporting via e-mail was possible in 2 services.

ARE REPORTING MECHANISMS AGE-APPROPRIATE, USER FRIENDLY AND ACCESIBLE AT ALL TIMES?

As part of this study, different testing scenarios were designed depending on the reporting mechanisms available on the service being tested. In the cases where more than one reporting mechanisms were available, researchers were instructed to test the reporting mechanism that they thought was most likely to be used by a 15 year old to report the situation created for the test. Aspects such as the prominence and the accessibility of the reporting mechanism were taken into consideration, but also if users' reports were acknowledged and responded to promptly.

One of the tests consisted in setting up a realistic bullying situation. A (fake) 15-year old girl user reported that she had been harassed on that particular service. The scenario consisted of one minor being bullied by two other minors who posted nasty comments on the profile of the "victim" and who uploaded some hurtful pictures. The "victim" reported the situation to the provider. If at all possible a message was added to the report. This message was carefully designed and worded to be a general request but at the same time a cry for help. Other testing scenarios consisted in asking the service provider to remove (or provide help to remove) some inappropriate content including photos, videos or comments depending on the possibilities offered by the platform being tested. Testing on the website revealed that **all the services assessed provide age-appropriate, user-friendly and easily accessible reporting mechanisms.** The reporting mechanisms were at all times available.

HOW ARE USERS REPORTS DEALT WITH?

In 8 services (including the one that was tested in two language versions) reports were acknowledged by the provider. In most cases, this acknowledgement consisted of an automated e-mail indicating how reports were usually handled and indicating a timescale for action. **7 services responded to the user reports** taking some action such as deleting or age-restricting the flagged content from the site, issuing warnings to the users who had violated the Terms of Use or guiding the reporting user on how to solve the problem themselves. In some cases this action was accompanied by a written reply to the user who had filed the report detailing what actions had been taken. **In 6 services, the responses were responded within 24 hours**. In one service a reply to the user was sent within 24 hours, but the reported content was deleted from the site only after 48 hours. The pie chart below (see Fig. 11) illustrates the response time to users who asked the Social Networking Services for help. .

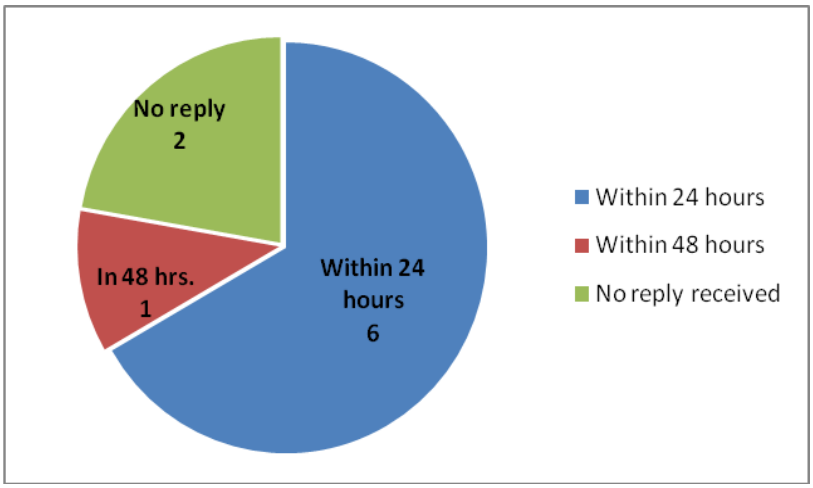

Fig. 11 Response time to user reports asking the Social Networking Services for help

Apart from the fact that the provider sent a reply to the (reporting) minor, attention was also paid to what actually happened with the offending content being reported (e.g. bullying pictures or comments posted on the profile of a "victim", flagged pictures or videos containing content considered as inappropriate for minors, etc.) and also to the actions, if any, taken against the offenders. **In 6 cases the inappropriate reported content was removed from the site or was age-restricted by the service provide**r and **only one service took some kind of action against the offender** (I.e. the "bully" created for this test). In this case, the user who had been reported received an email from the service provider reminding them of the Terms of Service and behaviours which constituted violations. They were informed that the provider felt that these had been violated by the user, and that they should cease any related activity immediately. The reported comments and photos were not automatically removed, but the user who had been reported was clearly informed that they should immediately delete content or activity from any parts of the account that were involved in the violation. However, they did not subsequently remove the bullying comments or pictures, and did not receive any further warnings as a result. In another service, the reported content was not automatically removed from the site, but the "victim" was guided on how to flag the inappropriate content on the "bullies" pages. The minor flagged the content. This time the report was responded and the offending content was removed from the site within 24 hours. In another test, the user report was acknowledged, the minor got no reply, but the offending message was deleted from the site within 24 hours.

As these results illustrate, the ways the different service providers respond to user reports vary considerably. Most services acknowledge reports, a few do not. Most services delete or age-restrict the reported content, while a few guide the users to do it themselves. Some services inform the users of the actions taken (e.g. offending content was deleted), while others delete the content but do not inform users that they did. In only one case, the offenders were warned, but even then, the reporting mechanism proved not completely efficient.

Excluding the new comers, findings regarding this principle are similar to the ones of last year's assessment.

EXAMPLES OF BEST PRACTICE FOR PRINCIPLE 4:

Flickr, Habbo Hotel, Skyrock, Stardoll and Windows Live are examples of best practice where user reports were acknowledged, the user reports were responded to quickly and an effective action was taken by the provider (e.g. the reported content was promptly removed from the site.)

Principle 5: Respond to notifications of illegal content or conduct

Because of ethical reasons Principle 5 was not tested on the website. Thus, this section only summarizes the main findings related to the analysis of the 9 self-declaration statements.

Principle 5 states that *"*upon receipt of notification of alleged illegal content or conduct[10]providers should have effective processes in place to expeditiously review and remove offending content." Principle 5 also states that "service providers should have in place arrangements to share reports of illegal content or conduct with the relevant law enforcement bodies and/or hotlines" (Safer Social Networking Principles for the EU, 2009: 8). While efficient processes for handling such notifications should be in place, it is evident from the analysis of the self-declarations that the nature of the measures implemented by each service varies according to the national legislation and jurisdiction where the services operate.

Principle 5 was evaluated as very satisfactory in all the services in relation to the self-declaration statements. In their self-declaration all the service providers claim that they have effective and expeditious processes in place to review (and eventually remove) content found offending from their services.

All the service providers assessed claim to have arrangements in place to share reports of illegal content with the corresponding law enforcement bodies. Only 3 services explicitly state in their self-declarations that they provide links on their websites to other local agencies and organisations in order to support the reporting of illegal content or conduct on their services.

Principle 6: Enable and encourage users to employ a safe approach to personal information and privacy

Principle 6 requires that the Social Networks *"Enable and encourage users to employ a safe approach to personal information and privacy".* Specifically providers should:

- Provide a range of privacy setting options with supporting information that encourages users to make informed decisions about the information they post online. These options should be prominent in the user experience and accessible at all times.
- Consider the implications of automatically mapping information provided during registration onto profiles, make users aware when this happens, and should consider allowing them to edit and make public/private that information where appropriate.
- Users should be able to view their privacy status or settings at any given time. Where possible, the user`s privacy settings should be visible at all times.

According to their self-declaration, Principle 6 was evaluated as *very satisfactory* in 4 services and as *rather satisfactory* in 5 services. In relation to the implementation of the self-declaration on the respective services,

[10] In the context of child protection, illegal content and conduct in this context refers to child abuse images and grooming respectively.

Principle 6 was assessed as *very satisfactory* in 7 services and as *rather satisfactory* in 2 of the services analysed (See Fig. 12).

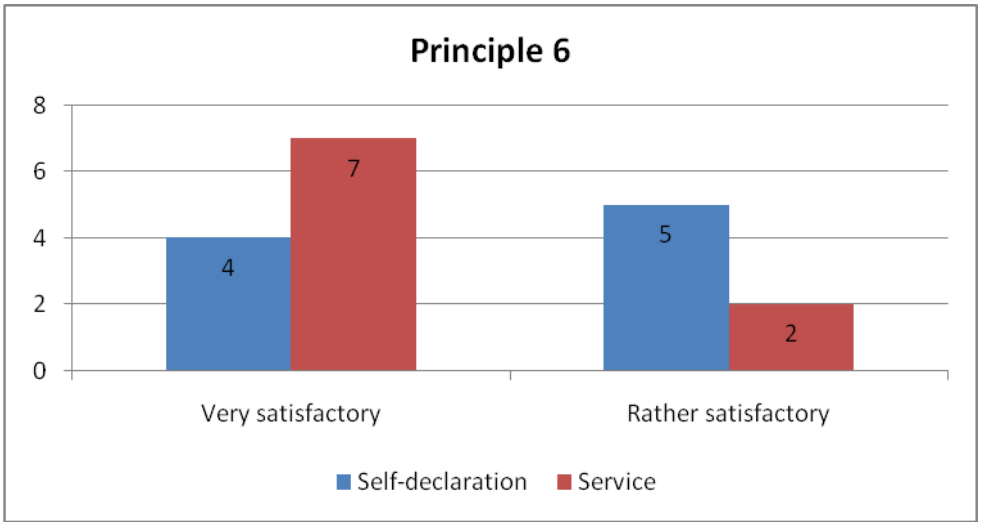

Fig. 12 Assessment of Principle 6 in self-declaration vs. service

WHAT TYPES OF PRIVACY SETTING OPTIONS ARE OFFERED TO USERS? ARE THESE SETTINGS USER-FRIENDLY AND ACCESSIBLE AT ALL TIMES?

All the services assessed offer their users (including minors) a range of privacy settings. These settings enable users to control who can have access to the information contained in their profile. **Privacy settings are user-friendly and accessible at all times in all the services analysed.**

Regarding the content posted on one's profile, some services offer users the option to set up privacy settings for individual pieces of content (e.g. upload videos or pictures as "private" or to allow only specific individuals to have access to such pieces of information). This option can usually be chosen during the content upload, but this can also be modified at any time after publication. Some services provide privacy settings that allow distinctions between different types of users, for instance, "contacts", "friends and family", "VIP friends", "all users", etc. and can be used to allow or restrict access to certain types of content or activities (e.g., deciding who can have access to guestbook, photos, videos or blogs; tagging pictures or videos; commenting on someone's profile, etc.) for different users. Privacy settings also include the ability to block other users and contacts. This may prevent access to a user's content or profile of the user who has blocked them, and/or use of the internal search function to locate them. Privacy settings also include options such as the possibility to restrict interactions among users, for instance by setting one's account up so that the minor can only interact with specific users or with nobody (e.g. registered users, all friends, a selected group of friends, nobody, etc.).

The range and modality of privacy options vary from service to service. Some services offer advanced, but still user-friendly privacy options that allow users to decide which individual piece(s) of content to share with specific (groups of) users. Services which offer more advanced settings allow users to modify individual settings with a finer degree of control over which aspects of a user's information are visible and the kinds of online communication that are allowed. Other services offer more limited privacy settings which lack complexity and do not allow users to customize privacy settings regarding specific groups of people or specific content.

ADDITIONAL PRIVACY SETTINGS AND PARENTAL CONTROLS

In one service which offered optional parental controls the privacy settings for the account were tested both with and without the use of parental controls. When used without parental controls, three main types of settings 'Private', 'Limited' and 'Public' were available. Even though by default, privacy settings for minors are

set to 'Limited'[11], greater levels of privacy are enabled by default using the parental controls facility. Contact management is also set by default to be controlled by parents who can then choose which communications services to allow and who the child can communicate with. The default setting may be changed to allow the child to manage their own contacts but this still allows the parent to monitor the contact list.

In another service tested, default settings enable maximum privacy through standard profile options for child and teen accounts. Modifications may be made which alter specific aspects of information visibility. However, teens and child users require parental consent to make any changes. A parent must log in to the service to approve any changes to safety settings. By default, access to a child account profile is blocked while a teen account profile is visible to friends only.

IS SUPPORTING INFORMATION REGARDING PRIVACY AND PRIVACY SETTINGS PROVIDED?

Privacy settings options are accompanied by supporting information in all the services assessed. Supporting information and guidance help users make informed decisions regarding their privacy settings options and/or instruct users on how to change their privacy settings. Depending on the service, supporting information may be included in one or several sections of the site, for instance in the specific page containing the privacy settings, in the Account Management panel, in the Community Guidelines, in FAQs and/or in Safety Guides provided on the service itself. Supporting information usually refer to both technical information on how to manage privacy settings and guidance about disclosing personal information on the service. In some services, users are informed when they first set up an account and reminded in help articles about never giving out personal information such as real name, address, email, or telephone number. Other services advise users to enable additional settings to limit access for certain information to 'close friends' only and provide online resources with additional guidance on the importance of privacy controls. Contextual information explaining how to modify specific privacy settings at the moment user-generated content is uploaded or updated is also provided in a few services.

IS INFORMATION PROVIDED BY MINORS DURING REGISTRATION AUTOMATICALLY MAPPED INTO THEIR PROFILES?

**In 5 of the 9 services assessed (some of) the information provided by minors during registration is automatically mapped into the user`s profiles.** However, only in 3 cases users were made aware that this would happen. Still, in 4 of the services that automatically map information into the user`s profiles it was possible to make (some or all of) this information private or public at a later stage. In one service, optional personal information (e.g. home address, favorite music, etc.) was not required in order to sign up, but may be added to one`s profile if the user wishes to do so. If a user decides to add extra personal information, this information is automatically mapped onto the user's profile and it becomes public. However, users are not explicitly informed of this and they cannot make this information private either. In another service, only some pieces of information (e.g. last name and date of birth) could be made public or private at all times. The rest of the information contained in users` profiles was either always public (e.g. gender) or always private (e.g. telephone number). It is made clear to users, though, which pieces of information are public and which are not. In another service, some automatic mapping of information collected on registration takes place, but privacy settings allow for control of visibility of information displayed. User content such as the profile guestbook and blog are by default open to all to post, but may be made completely private or restricted to friends only. In this service, it is not possible to restrict a full profile to either 'friends only' or to nominated friends. Some information and functions may be restricted but all other aspects of a user's profile remain visible to all and for this reason ensuring that profiles do not contain any personal information is essential. Only in 4 services the personal information required during registration is not automatically mapped onto the user profile.

[11] By default, in this service, privacy settings for minors are set to 'Limited' i.e. the user's profile containing any descriptive information about general interests, occupation and location is visible to all, but other information such as status, contact information and access to photo albums is restricted to friends only

IS IT EASY TO DELETE ONE`S PROFILE ON THE SNS?

In 7 of the 9 services tested it was possible for users to delete their profile themselves. One service does not delete profiles on request. However the FAQ advices users who wish to delete their profile to stop using their account, delete all the items belonging to the user account and spend all credits that may be over because the administration removes accounts from the database that are not used in over 12 months as part of routine service maintenance. The suggested alternative though may be seen as difficult or overly time consuming by children and young people. In 1 service it is not possible to delete a profile unless the main ID account (which is also used to access all the other services offered by this provider such as e-mail, photo-sharing applications, etc.) is terminated. In the latter case, though, the privacy settings can be used to hide the profile from all other users.

Deletion of a user profile is straightforward and easy to manage in the 7 services that provide this option. When deleting one`s profile 6 services state what would happen to the user's personal information and user-generated content contained in their profile. For instance, they communicate users if their personal information will be completely deleted or if (some of) this information will be retained. If information is retained they usually refer to how this information may be used after the profile is deleted.

EXAMPLES OF BEST PRACTICE FOR PRINCIPLE 6:

Flickr, Habbo Hotel, Yahoo! Pulse, Windows Live, Xbox and YouTube are best practice examples of websites that offer advanced privacy settings that allow users to have a finer degree of control over which aspects of a user's information are visible and the kinds of online communication that are allowed on these services.

Principle 7: Assess the means for reviewing illegal or prohibited content/ conduct

As with Principle 5, because Principle 7 was not tested on the website, only the main findings of the analysis of the self-declarations are summarised here.

Principle 7 has to do with the reviewing of illegal or prohibited material on the Social Networking Sites. According to it, the SNS providers should *"during the normal course of developing and managing SNSs, assess their service to identify potential risks to children and young people in order to determine appropriate procedures for reviewing reports of images, videos and text that may contain illegal and inappropriate/unacceptable/prohibited content and/or conduct". S*uch procedures include ("Safer Social Networking Principles of the EU," 2009: 9):

- human and/or automated forms of moderation
- technical tools (e.g. filters) to flag potentially illegal or prohibited content
- community alerts
- user-generated reports

In addition, Principle 7 states that providers who employ human moderators who interact in real-time with children and young people should take reasonable steps "to minimise the risk of employing candidates who may be unsuitable for work which involves real-time contact with children or young people" ("Safer Social Networking Principles of the EU," 2009: 9).

According to the analysis of the self-declarations, Principle 7 was assessed as *very satisfactory* in all the services assessed. The main mechanisms for services to identify potential risks to minors are user-generated reports and the use of automated mechanisms such as employing text or image filters to identify potential threats to minors. According to the 9 self-declaration statements analysed, both mechanisms are available in all the services tested.

According to the self-declarations, specially trained personnel reviews inappropriate content in all the services analysed. Three service providers state that they employ human moderators who interact in real-time with children and young people. However, only in 2 of these services the self-declaration explicitly refers to the steps taken to minimize the risk of employing candidates who may be inappropriate for interacting with minors in real-time. These include strict selection procedures prior to hiring moderators and dedicated training specially tailored for them.

CONCLUSIONS

The assessment of Principle 1 is very positive on the website and consistent with the providers` commitments in their self-declarations. In terms of the availability and easiness of the safety information on the websites, all the services tested in this Phase include in their services safety information, guidance and/or educational materials specifically targeted at children, their parents or guardians and teachers. In all these services this information was easy for minors to understand and in 7 services it was also easy to find. Excluding the newcomers, findings regarding this principle are similar to the ones of last year's assessment.

Principle 2 was assessed as *very satisfactory* in 7 services and as *rather satisfactory* in 2 services. The best evaluated services were those where access was denied to underage users, where effective mechanisms to prevent re-registration were in place and where minors were not confronted with any type of inappropriate content. Nevertheless, as no 100% reliable age-verification mechanism exists up to date, even in the cases where age-restriction mechanisms proved to be effective, it was possible for underage users to register on the site by creating a completely new user account with an age above the minimum required by the service. Regarding the appropriateness of content found on the services, this depends largely on the adequate tagging or labelling of the material uploaded by users themselves. This is specially the case of photo-sharing and video-sharing platforms where large amounts of content are uploaded every minute. During testing labelling or tagging mechanisms (associated to internal filtering processes) proved to be effective, although not infallible because in a few cases it was possible to identify content that could be considered as inappropriate for younger children or adolescents.

As regards Principle 3, results show that in only one of the services tested profiles of minor could be found in internal and external search engines by name searches. Nevertheless, in the other services tested the profiles of minors could be found via other mechanisms such as searches for user-generated content or via friends' profiles. For instance, once a minor's blog, individual videos, pictures, video channels or photostream are found (e.g. via an external search engine), access to the minor's profile is granted. Furthermore, only 2 services make minors' personal and identifiable information visible by default only to their approved list of contacts. This means that even if minors add extra information about themselves (not required during registration) to their profiles this information is not displayed to users beyond the minor's contacts by default. In 6 services a considerable amount of personal information was displayed to users beyond the minor`s approved contacts list. In one service limited personal information was displayed, but it was accessible to friends of friends. It is important to stress, however, that this information did not always lead to the identification of minors and, therefore, it does not necessarily constitute an imminent safety risk. In addition, platforms tested in Phase B are not "typical" social networking sites. This means that in several services user profiles are not the main point of entry or of interest for a user. This is specially the case of photo-sharing and video-sharing platforms where users would typically visit a photostream or a video channel, but not necessarily a user's profile. Still, by default, in all the services tested minors could be contacted via friend requests, in 7 services they could be contacted via public messages and in 3 services they could be contacted via private messages by users beyond their approved list of contacts.

In relation to Principle 4, testing indicated that most services assessed in this Phase acknowledge user reports and delete or age-restrict the reported content, while a few guide the users to do it themselves. In only one case, though, the offenders received a warning for having violated the Terms of Service. Excluding the new comers, findings regarding this principle are similar to the ones of last year's assessment.

Concerning Principle 6, all the services assessed offer their users (including minors) a range of privacy settings. These settings enable users to control who can have access to the information contained in their profile. Privacy settings are user-friendly and accessible at all times in all the services analysed. However, a few services offer users (including minors) a limited set of privacy options which lack complexity and do not allow users to customize privacy settings regarding specific groups of people or specific content.

REFERENCES

European Commission. (2009). Safer social networking: The choice of self-regulation. Retrieved 02.08.10, from http://ec.europa.eu/information_society/activities/social_networking/eu_action/selfreg/index_en.htm#self_decl

Nielsen, J. (1990). How to conduct a heuristic evaluation. Retrieved 03.03.10, from http://www.useit.com/papers/heuristic/heuristic_evaluation.html

Safer Social Networking Principles for the EU. (2009, February 10th). Retrieved 02.08.10, from http://ec.europa.eu/information_society/activities/social_networking/docs/sn_principles.pdf

Staksrud, E. and Lobe, B. (2010). *Evaluation of the implementation of the Safer Social Networking Principles for the EU Part I: General Report.* European Commission Safer Internet Programme, Luxembourg.

FOR FURTHER INFORMATION:
DIRECTORATE-GENERAL
INFORMATION SOCIETY AND MEDIA
EUROPEAN COMMISSION
SAFER INTERNET PROGRAMME
E-MAIL:
SAFERINTERNET@EC.EUROPA.EU
FAX: + 4301 34079
OFFICE: EUFO 1194
EUROPEAN COMMISSION
L-2920 LUXEMBOURG

http://ec.europa.eu/saferinternet